|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  9th Meeting: Geneva, CH, 27 April – 7 May 2012 | Document: JCTVC-I0472 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Considerations for the creation of HEVC profiles and levels** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Requirements for consideration | | |
| *Author(s) or Contact(s):*  *Other supporters of this document:* | Mukta Kar Cable Television Laboratories (CableLabs)  858 Coal Creek Circle  Louisville, Colorado, USA 80027  Arianne T. Hinds  Cable Television Laboratories (CableLabs)  858 Coal Creek Circle  Louisville, Colorado, USA 80027  Yasser F. Syed  Comcast Cable/ Comcast Labs  4100 E. Dry Creek Rd  Centennial, Co 80122  Charles Hasek  Time Warner Cable  Clint Sheridan  Rogers Communication of Canada  Jason Barton  Cox Communications  William Warga  Liberty Global, Inc.  Steve Maxwell  Brighthouse Cable  Mark Francisco/Yasser Syed  Comcast Cable  David Colter  Charter Communications  Glenn Reitmeier  NBC-Universal | Tel: Email:  Tel:  Email:  Tel:  Email:  Email:  Email:  Email:  Email:  Email:  Email:  Email:  Email: | +1.303.661.3842 [m.kar@cablelabs.com](mailto:m.kar@cablelabs.com)  +1.303.661.3419  [a.hinds@cablelabs.com](mailto:a.hinds@cablelabs.com)  +1.303.246.8413  [yasser\_syed@cable.comcast.com](mailto:yasser_syed@cable.comcast.com)  [charles.hasek@twcable.com](mailto:charles.hasek@twcable.com)  [clint.sheridan@rci.rogers.com](mailto:clint.sheridan@rci.rogers.com)  [jason.barton@cox.com](mailto:jason.barton@cox.com)  [bwarga@lgi.com](mailto:bwarga@lgi.com)  steve.maxwell@mybrighthouse.com  [mark\_francisco@cable.comcast.com](mailto:mark_francisco@cable.comcast.com)  Yasser\_syed@cable.comcast.com  david.colter@chartercom.com  Glenn.Reitmeier@nbcuni.com |
| *Source:* | Members of AHG8 for the study of profiles and levels | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This document summarizes use cases and offers requirements for consideration in the creation of HEVC profiles that can be regarded as relevant to the applications and services provided by Multiple Systems Operators (MSOs).

# Introduction

MSOs are faced with the demand for services and support of applications at different service rates, ranging from traditional broadcast of standard and high definition content to the support of mobile devices, video on demand, adaptive bit rate streaming, IPTV, networked personal video recorders, home monitoring and surveillance, and home network video distribution. Applications envisaged for the near term further include video teleconferencing, and support of second screens. Some of these applications may coincide on a single device. The devices typically employed in these current and future scenarios include: set top boxes utilizing a QAM interface or connected via IP, tablets, smart phones (both with large and small screens), and personal computers. Amongst these devices, underlying chip sets may include single-core or multi-core processors; memory constraints vary, and the power source may be a wall outlet or a battery, while at the same time, monitored in support of green energy initiatives. Stream encoding may be optimized for high visual quality (e.g. for premium channels, 4Kx2K displays) or optimized for bitrate efficiency to satisfy bandwidth management constraints. To address these various scenarios efficiently, only a minimal number of profiles should be defined, but enough so as to not create “unofficial or constrained” profiles that are created to minimize decoder costs for the application/device. For example, corner cases that may spur the development of unofficial profiles include conforming streams whose performance parameters exceed the intended use of application(s) tools and that significantly increase decoder complexity, or the questionable combination of applications that would not typically exist within a single product (e.g. contribution vs. delivery services). On the other hand, a large number of profiles, each targeting a specific application or device could result in slower rates of adoption and potential problems with interoperability; hence integration, where possible, is encouraged. The profile(s) need(s) to be sufficient to meet the various categories of requirements for all scenarios across the processing and power-source capabilities of these devices. It may be that a single device is capable of supporting one or more profiles depending on the application(s) and that the development of a candidate HEVC profile to support applications residing on a single device should be considered, but the profile should not overly burden decoders with increased implementation costs for support of applications or requirements for performance that are not likely to coincide on a single device. The discussions of profiles should minimize the number of them to preserve interoperability and facilitate adoption, and is therefore a suggested next step for the JCT-VC committee.

It is also believed that profiles should not incur profile “obsolesce”; that is a newer profile which replaces the defined profile. This may be due to new revised functions whose capabilities were not fully realised till later investigations. Profiles should be defined such that any new profile doesn’t replace an older profile, but stands on its own and enables a different class of applications. If replacement does occur, then cost and confusion will increase due to existing product replacement while confidence will be reduced.

One possible approach for defining HEVC profiles is to identify and group the major application scenarios (along with their targeted devices and subsets of requirements), and fit a profile to each major group, while simultaneously minimizing the number of profiles. For example, the committee may consider a profile to maximize coding efficiency and visual quality for applications employed on set top boxes or other devices with multi-core processors, and another profile for devices where both bitrates and complexity are reduced because of limited availability of computational resources or power.

Table I provides a summary of applications, associated devices, and typical requirements supported or expected by MSOs, and is intended as a data point to aid in the definition of major application groups for which possible HEVC profiles may be developed.

**Table I. Summary of typical requirements, devices, and applications supported by MSOs.**

|  |  |  |  |
| --- | --- | --- | --- |
| Application  category | Application description | Devices | Typical Requirements |
| 1 | Broadcast video including video on demand, network personal video recorders, and live video | * Set top boxes * Digital transmission adaptors up to high definition resolution | * Resolutions including: 1080P@24,30,60 fps, , 4Kx2K@24,30,60 fps, 1080i30/720P24,60, and [720x480i@29.97](mailto:720x480i@29.97), [720x480p@59.94](mailto:720x480p@59.94) Resolutions * Support of 120 fps * Fixed frame-rate mode * Low delay support in fixed frame-rate mode * Constant quality content through multi-pass encodings using variable bitrate encoding. * Indexed I-Frames or sub-GOP for dynamic trick-play support at 2-3 speeds. Frame decoding for trick-play is dependent only on a subset of decoded frames in the stream. * Higher bitrate with correlated pre-processing filters (possibly with higher complexity) for higher quality for premium content channels, e.g. encoder prioritizes high quality] * Lower bitrate and lower quality for lower tier channels, e.g. encoder prioritizes bit rate. |
| 2 | HTTP based adaptive bitrate streaming and internet protocol television. | * Tablets * Mobile phones * Personal computers * Second screen mobile devices * IP set top boxes * Other devices that may act as portable high quality storage device * Decoders that may display on a TV monitor (e.g. via HDMI) | * IP set top boxes with the same requirements as application category 1. * Higher quality (at resolutions up to 1080P) at 60 fps for tablets and mobile devices. * Encode Content for Multiple Bit Rate streams that align at conditioned points |
| 3 | Video conferencing | * Tablets * Personal computers * TVs used as monitors * STBs | * SD, QHD, and HD resolutions * Bi-directional encoding of live content. * 1080i/720P HD resolutions at minimum. * Low latency transmissions |
| 4 | Home Monitoring/ Surveillance | * Set top boxes/TVs * Tablets * Personal computers | * SD and HD resolutions * Time lapse scanning * Buffer scanning |
| 5 | Wireless network in-house and edge video distribution | * Edge devices including personal computers, tablets, smartphone and other multi-core devices | * Storage and transcoding in edge devices for on-demand services May come in as AVC or HEVC * stored or transcoded to required HEVC profile and bitrate * On demand in in-house device or edge device and vice versa as well. * Want to lower the probability of catastrophic failure on wireless transmissions. |
| 6 | Multi-Service encoders and transcoders | * set top boxes * single-core, small screen smartphones * multi-core devices * mobile/second screen devices | * Common set of resolutions and bit rates that can be adapted into multiple services. * Reuse of streams from ABR set for multiple services. * Higher bitrate streams on set top boxes * Lower bitrate streams on second screen devices * Lowest bitrate streams used on limited devices including single-core, small screen smartphones. * Concept of mezzanine encoding. |
| 7 | Broadcast of video adapted to mobile devices. | * Tablets * Smart phones | * Need transcoding and the ability to wrap into different containers. |
| 8 | Contribution | * Production level devices such as Contribution encoders, Editors, Switchers, etc. | * 4:2:0 – 8 bit * 4:2:0 – 10 bit * 4:2:2 – 10bit * 1080i30, 1080p60, 720p60 * 4k x 2k@30p, 60p, 120p |

## Additional considerations for HEVC profiles

While the above table serves to define major application categories for MSOs, the following list suggests criteria that could further aid in the development of multiple profiles for HEVC:

* Candidate streams should be decodable by both multi-core and single core devices
* Products targeting applications requiring contribution/mezzanine streams should be separable from products targeting applications where premium or standard service delivery quality streams are used.
* The coding and visual correction tools required in a profile for a particular group should not excessively add to decoder complexity. A clear delineation between profiles will also serve to prevent the creation of informal, constrained profiles. Implementation complexity for decoders supporting standard quality streams should not significantly exceed twice the complexity of decoders targeted for AVC level 4.0 ranges.
* Profiles should provision a mode that can simplify stream processing. For instance fixed frame rate can simplify timing, HRD modeling, and low delay modes. It should be maintained that when low\_delay\_hrd\_flag is not present then its value is inferred to be equal to 1-fixed\_pic\_rate\_flag.
* There is a significant amount of interlaced content that exists and continues to be created; all of which will need to be encoded in HEVC to enable complete video services for MSO customers. While it is clear that there is a trade-off between complexity and granularity for interlacing tools, to best realize the coding efficiencies and features of HEVC, the (relatively simple approach of) signalling interlace content may not be sufficient. For example, some tools should be able to operate on frame/field level without a large increase in complexity, and may even result in reducing the size of coded progressive I pictures. If these tools are discovered to be useful in the encoding of multiple types of content (including interlace content), then a new profile that employs them may emerge, further adding to the likelihood of profile obsolesce and the proliferation of profiles. Based on these concerns, there should be more focus placed on providing support for interlaced formats beyond the signalling of interlace content.

It is proposed that an investigation is conducted to determine the quality state of interlacing content as compared to progressive content. If it is determined that more optimized interlacing tools are needed, then it is proposed to have a separate profile that includes these interlaced tools. Additionally it is proposed to add ammedments to reflects these new tools as well.

* The support of trick-mode behaviour should also be an area of focus to avoid profile obsolesce and to manage the implementation complexities in the decoder for the support of trick-modes. In the past, the support of multiple trick speeds was accomplished by the creation of separate trick-files, which resulted in an extra burden in decoders, i.e. decoders needed to process more frames than what was required for display at a particular trick-play speed. By creating/bounding a sub-list of reference pictures from the start of a RAP, decoding processors can avoid becoming overdriven for the support of multiple trick-mode speeds. For example, to enable less burdensome trick-modes by extracting decidable sub-sequences, each picture must not require processing or the decoding of any picture with a higher temporal\_id values. To accomplish this, the reference indices could comply to the following order: 1) Each reference picture with a temporal\_id value equal to or less than the value of the temporal\_id of the current picture shall have an index value that is less than the respective index values corresponding to each reference picture with a higher temporal\_id value; 2) Non-reference pictures have the highest temporal\_id value or the highest temporal\_id value minus one; 3) IDR & CRA pictures shall have the same temporal\_id value of the lowest value; 4) All other non IDR or CRA value must have a higher temporal\_id value than the IDR or CRA.
* Clearly defined profiles and levels will help to prevent confusion. Overlapping terms already used for AVC/H.264 should be minimized. For example, use of the term “Main profile” could be changed to another term, perhaps “Central profile” to avoid confusion with AVC.
* New profiles, if defined, should enable functionalities and applications that target performance targets that are higher than the currently considered Main Profile.

## Considerations for HEVC Levels

Decoder costs can be increased by excessive maximum decoder performance bounds specified at a target level. Some class of applications may not even approach this limit (broadcast/delivery), though other applications (I Frame editing) may be around this level. Costs need to be contained for services especially decoder costs which need to be minimized when possible. Applications planned today span SD, HD, and 4K resolutions, but even levels supporting 4K will have applications such as broadcast/delivery that may not even approach performance bounds of the level below.

Table A-1 (appendix) in the current draft text specifies level limits in a hierarchical structure with consecutive levels supporting the same parameters with the only differences being maximum bit rate and maximum CPB size.

Table 2 (appendix) is provided based on current discussions, with some additional items added and highlighted to illustrate the following points:

* Levels 4 and 4.1 support the same parameters (including resolutions) with the only differences being maximum bit rate and maximum CPB size, 15 Mbps and 30 Mbps, respectively.
* Levels 4.2 and 4.3 support the same parameters (including resolutions) with the only differences being maximum bit rate and maximum CPB size, 30 Mbps and 50 Mbps, respectively.
* Levels 5 and 5.1 support the same parameters (including resolutions) with the only differences being maximum bit rate and maximum CPB size, 50 Mbps and 100 Mbps, respectively.

It is assumed that the intent is to allow applications to choose levels based on whether they have “lower” bit rate/CPB size requirements or “higher” bit rate/CPB size requirements.  This contribution recognizes the value of defining two sets of levels to address these different requirements.  For example, certain applications such as broadcast requirements may find the “lower” requirements sufficient for their needs while avoiding the cost associated with supporting the “higher” requirements while other applications such as content distribution may require the “higher” requirements for their needs.

However, attempting to satisfy both these application types using a single consistent set of level limits is difficult and places unnecessary constraints on each application type.  For example,

* Level 4.2 support (the “lower” level for 1080p60) includes support for Level 4.1 (the “higher” level for 720p60 requiring 30 max Mbps) and not strictly support for Level 4 (the “lower” level for 720p60 requiring 15 max Mbps).
  + This forces Level 4.2 support to be higher than 30 Mbps to be consistent with level 4.1 (or can be thought as forcing Level 4.1 to not be higher to ensure Level 4.2 support is consistent) when different values may be more appropriate for the level definitions.
* Level 5 support (the “lower” level for 4Kp30) includes support for Level 4.3 (the “higher” level for 1080p60 requiring 50 max Mbps) and not strictly support for Level 4.2 (the “lower” level for 1080p60 requiring 30 max Mbps).
  + This forces Level 5 support to be higher than 50 Mbps to be consistent with level 4.3 (or can be thought as forcing Level 4.3 to not be higher to ensure Level 5 support is consistent) when different values may be more appropriate for the level definitions.

It is felt that the level definitions especially at higher levels needs to be investigated and aligned with classes of applications. Being cognizant of application uses during level definitions of the profile(s) will lead to more cost optimized decoders for those reduced demanded applications.

There can be several approaches that can be used, but only a few that can clearly help in defining product based on profile/level definition and class of application. Therefore, it is recommended for discussion that the following changes be implemented to define levels more suitable to application types that have “lower” bit rate/CPB size requirements and those that have “higher” bit rate/CPB size requirements:

* Define two clearly separate sets of levels with different naming conventions that avoid confusion between the two sets.  For example, Levels 1.0, 2.0, 3.0, etc. in one table and Levels High1.0, High2.0, High3.0 in another table.
* Define separate profiles (even though we recognize that profiles are generally used to distinguish coding tools and not simply parameter limits) to more clearly delineate the differences.

# References

[1] Text of Committee Draft ISO/IEC 23008-2, High Efficiency Video Coding, w12556; JCT-VC H1003, High Efficiency Video Coding (HEVC) text specification draft 6; February 2012, San Jose, CA, USA.

# Patent rights declarations (s)

Cable Television Laboratories and Comcast Cable/Comcast Labs do not have any current or pending patent rights relating to the technology described in this contribution.

# APPENDIX A

Table A-1 in the current draft text [1] specifies level limits in a hierarchical structure with consecutive levels supporting the same parameters with the only differences being maximum bit rate and maximum CPB size.

**Table A-1: replicated from [1] and provided for reference.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Level** | **Max luma pixel rate MaxLumaPR**  **(samples/sec)** | **Max luma picture size MaxLumaFS (samples)** | **Max bit rate MaxBR**  **(1000 bits/s)** | **Min Compression Ratio MinCR** | **MaxDpbSize (picture storage buffers)** | **Max CPB size**  **(1000 bits)** |
| **1** | 552,960 | 36,864 | 128 | 2 | 6 | 350 |
| **2** | 3,686,400 | 122,880 | 1,000 | 2 | 6 | 1,000 |
| **3** | 13,762,560 | 458,752 | 5,000 | 2 | 6 | 5,000 |
| **3.1** | 33,177,600 | 983,040 | 9,000 | 2 | 6 | 9,000 |
| **4** | 62,668,800 | 2,088,960 | 15,000 | 4 | 6 | 15,000 |
| **4.1** | 62,668,800 | 2,088,960 | 30,000 | 4 | 6 | 30,000 |
| **4.2** | 133,693,440 | 2,228,224 | 30,000 | 4 | 6 | 30,000 |
| **4.3** | 133,693,440 | 2,228,224 | 50,000 | 4 | 6 | 50,000 |
| **5** | 267,386,880 | 8,912,896 | 50,000 | 6 | 6 | 50,000 |
| **5.1** | 267,386,880 | 8,912,896 | 100,000 | 8 | 6 | 100,000 |
| **5.2** | 534,773,760 | 8,912,896 | 150,000 | 8 | 6 | 150,000 |
| **6** | 1,002,700,800 | 33,423,360 | 300,000 | 8 | 6 | 300,000 |
| **6.1** | 2,005,401,600 | 33,423,360 | 500,000 | 8 | 6 | 500,000 |
| **6.2** | 4,010,803,200 | 33,423,360 | 800,000 | 6 | 6 | 800,000 |

Table 2 is a modified version of table A-1, with the following changes made: a column has been added to reflect the typical picture sizes from the parameters defined in table A-1, and an additional resolution for QHD (960x544@30p) has been added for Level 3. Note that the corresponding max/min parameters for Level 3 currently do not reflect support for QHD (they have yet to be modified). Levels 4, 4.1, 4.2, 4.3, 5, and 5.1 are highlighted to illustrate the points made in section 1.2.

**Table 2: Modified version of A-1 from [1] with content highlighted for points of discussion, QHD added to level 3.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **HEVC Working DraftLevel** | **Picture Resolution** | | **Max luma pixel rate MaxLumaPR**  **(samples/sec)** | **Max luma picture size MaxLumaFS (samples)** | **Max bit rate MaxBR**  **(1000 bits/s)** | **Min Compression Ratio MinCR** | **MaxDpbSize (picture storage buffers)** | **Max CPB size**  **(1000 bits)** |
| **1** | QCIF or below | 552,960 | | 36,864 | 128 | 2 | 6 | 350 |
| **2** | 352x240@30p  427x240@25p | 3,686,400 | | 122,880 | 1,000 | 2 | 6 | 1,000 |
| **2.x** | 640x360@30p\* | ? | | ? | ? | ? | ? | ? |
| **3** | 720x480@30p,  720x576@25p, 854x480@30p,  960x544@30p | 13,762,560 (?) | | 458,752 (?) | 5,000 | 2 | 6 | 5,000 |
| **3.1** | 960x544@60p  1280x720@30p | 33,177,600 | | 983,040 | 9,000 (10Mbps)\*\* | 2 | 6 | 9,000 |
| **4** | 1280x720@60p 1920x1088@30p | 62,668,800 | | 2,088,960 | 15,000 | 4 | 6 | 15,000 |
| **4.1** | 1280x720@60p, 1920x1088@30p | 62,668,800 | | 2,088,960 | 30,000 | 4 | 6 | 30,000 |
| **4.2** | 1920x1088@60p 2048x1088@60p | 133,693,440 | | 2,228,224 | 30,000 | 4 | 6 | 30,000 |
| **4.3** | 1920x1088@60p 2048x1088@60p | 133,693,440 | | 2,228,224 | 50,000 | 4 | 6 | 50,000 |
| **5** | 3840x2160@30p 4096x2160@30p | 267,386,880 | | 8,912,896 | 50,000 | 6 | 6 | 50,000 |
| **5.1** | 3840x2160@30p 4096x2160@30p | 267,386,880 | | 8,912,896 | 100,000 | 8 | 6 | 100,000 |
| **5.2** | 3840x2160@60p 4096x2160@60p | 534,773,760 | | 8,912,896 | 150,000 | 8 | 6 | 150,000 |
| **6** | 7680x4320@30p | 1,002,700,800 | | 33,423,360 | 300,000 | 8 | 6 | 300,000 |
| **6.1** | 7680x4320@60p | 2,005,401,600 | | 33,423,360 | 500,000 | 8 | 6 | 500,000 |
| **6.2** | 7680x4320@120p | 4,010,803,200 | | 33,423,360 | 800,000 | 6 | 6 | 800,000 |

\*This resolution is used for Adaptive Streaming/ATSC for mobile TV service in US.

\*\* To harmonize with Ultra Violet CFF